DOCUMENT RESUME

ED 138 780 CE 011 014

AUTHOR Dixon, Jane K.
TITLE Methodological Considerations in Evaluation of Continuing Education in the Health Professions--Utilization of Chain of Events Models in Determining Evaluation Criteria.
PUB DATE 4 Apr 77
NOTE 27p.; Paper presented at the Annual Convention of the American Educational Research Association (New York City, April 4, 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Educational Assessment; *Evaluation Criteria; Evaluation Methods; Evaluation Needs; Medical Care Evaluation; *Medical Education; Medicine; Nursing; *Professional Continuing Education; *Program Evaluation; *Research Methodology; Research Problems; Research Reviews (Publications)

ABSTRACT

An evaluation strategy is suggested which is designed to determine the effects of continuing education programs on the health care delivery system and to augment that effect. Topics dealt with include the distinctive aspects of continuing education; evaluation criteria; relationships among evaluation criteria; results of a review of literature on continuation education evaluation in medicine and nursing; limitations of evaluation; and conclusions regarding measurement, inference, and supporting research. The author concludes that if evaluation of continuing education programs is to be both successful and realistic, it cannot stand alone; it should be supplemented by two other lines of research: (1) All research into the effects of continuing education does not need to be of the "tacked on evaluation" variety. It is possible to create projects of which the primary purpose is to study continuing education, rather than to practice it. Such studies may serve as demonstration projects. (2) In general, to show that specific behavior inputs by health professionals are related to specific patient outcomes is the task, not of program evaluation, but of clinical practice research. Results of such experimental studies can be used by planners and evaluators of educational programs in drawing inferences about possible program results and in minimizing need for measurement in some areas. (LAS)

**********
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
**********

ED138780

# Methodological Considerations in Evaluation of Continuing Education in the Health Professions-- Utilization of Chain of Events Models in Determining Evaluation Criteria

Jane K. Dixon
Yale University
School of Nursing

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Paper presented at Annual Convention of the American Educational Research Association
New York City, April 4, 1977

OE 011 014

This is a revised version of the paper, Evaluation of Continuing Education Programs in Nursing -- Balancing Ideals and Realities, which was presented at the Oncology Nursing Program Evaluation Meeting, New Haven, Connecticut in January, 1977. That meeting was sponsered by the National Cancer Institute.

ABSTRACT

Though the ultimate goal of continuing education in the health fields is to bring about improvements in patient care through change in the work performance of clinicians, evaluations of such programs rarely involve collection of data which indicates directly whether this has been achieved. Instead, evaluations often focus on variables representive of lower level criteria (i.e., participants' perceptions, knowledge gain). In this paper the literature of continuation education evaluation in medicine and nursing is reviewed. An evaluation strategy designed to determine the effects of continuing education programs on the health care delivery system and to augment that effect is suggested.

There is a consensus that the ultimate goal of continuing education programs in the health fields is to bring about improvements in patient care through change in the behavior of practicing clinicians. Yet evaluations of continuing education programs in the health professions rarely involve collection of data which provides a direct indication of whether this goal has been achieved. In fact, variables representative of totally different constructs, such as participant satisfaction or knowledge gain, are frequently used as primary evaluation criteria. The purpose of this paper is threefold--1) to present a theoretical discussion in which the nature of continuing education in the health professions and possible criteria for its evaluation are explored, 2) to review the literature of continuing education evaluation in several predominant health fields in order to provide insight concerning the uses of the various criteria, and 3) to suggest a practical strategy for evaluation through which one can work toward achievement of ideal goals--the determination of the effect of continuing education programs on the health care delivery system and the augmentation of that effect--within the funding and operational restrictions commonly imposed on evaluation efforts.

The development of a productive, yet practical, model for evaluation of continuing education in the health professions is necessary for several reasons. 1) The pressure on practicing clinicians to engage in continuing education is on the increase, due, partially, to legislative mandates, policies and requirements of professional associations, and, in some fields, a tightening labor market. These extrinsic pressures can be fully justified only to the extent that there is solid evidence documenting the value of such study. To date, both the accumulation of such evidence in the form of published, empirically-based evaluation studies and the development of systematic methodologies for creating such evidence have lagged well behind the growth of continuing education programs.

2) The second factor necessitating the development of such an evaluation model relates the economic problems of our society, the financing of continuing education and a growing public skepticism regarding any and all social programs. The key concept which links these phenomena together is accountability. Though new ideas for educational programs, health programs and social programs are plentiful, both financial resources to be invested in such programs and public confidence in the value of such programs have, in the 1970's, become scarce commodities. Thus, programs must compete with each other for support, and the responsibility for demonstrating program effectiveness is placed squarely on the backs of program advocates. Belief systems are not enough; hard evidence is demanded.

3) The third factor is competing opportunities for continuing education within the health professions. Science, medicine, and related fields are in the midst of a knowledge explosion, and a clinician seeking to increase his informational base or update his skills has many alternative directions in which he may look--both in terms of established programs of study and in terms of potential content areas in which specialized study may be worthwhile. This situation increases the pressure on advocates and promoters of specific programs to demonstrate the effectiveness of those programs. Consumers now feel they have a right to be informed about the characteristics of the product with which they are becoming involved, and the adequacy of such information, as well as its content, may bear a crucial influence on the decision to become involved. This applies to the consumption of educational programs, as well as to consumption of material goods.

Distinctive Aspects of Continuing Education

In developing evaluation models for any specific type of education program, it is vital that the unique aspects of that type of program be carefully considered; evaluators of professional continuing education programs must avoid the trap of mechanically employing evaluation strategies developed for use in other types of educational programs. Characteristics which differentiate professional continuing education programs from non-professional programs and from programs providing basic professional education include the following; Program characteristics--specificity of focus, length of program, lack of control through grading, marginality to institution; Student characteristics--age, experience/competence, adult vs. student role, source of motivation, specificity of purpose, possibility for early implementation.

Student characteristics:

1) Age of students. It has been pointed out that most of our knowledge about the nature of learning is based on study of animals and of children (Knowles, 1973). The practice of andragogy, or the education of adults, is generally based on principles developed through study and practice of pedagogy, or the education of children. While this is inappropriate at the college or even the high school level, the inadequacy of pedagogy as a basis for educational practice decisions in continuing education is even more obvious.

2) Experience and competence of students. Students do not enter a continuing education with a blank slate; in fact, they may have extensive previous experience, even to the point of already being specialists in some aspects of the content covered by the course.

3) Lack of acceptance of student role. Experienced, mature people accustomed to being regarded as responsible adults do not easily adapt to the student role--to the extent that role carries overtones of dependency, childishness or passivity. Kramer (1974) has discussed in detail the "reality shock" which baccalaureate nursing graduates experience in moving from the role of student to that of worker. She reports that ideological differences between academic and service institutions are a major source of difficulty. Movement in the opposite direction may be equally troubling, particularly for a mature individual who is returning to an educational institution which differs widely from his expectations--expectations based on memories of a previous educational experience which occurred long ago and, perhaps, far away.

4) Source of motivation. The related processes of student evaluation, grading and granting of credentials seem to be primary sources of motivation for many students enrolled in degree programs, despite the protestations of their more idealistic faculty members that this should not be the case. While the external pressures on health professionals to engage in continuing education may serve a similar motivating function for some non-degree students, for others intrinsic sources of motivation derived from experiences as a practicing clinician may play a primary role. For example,

the thought of a patient who might have been helped more fully, awareness of a gap between discovered knowledge and his personal knowledge in an area, or an ambition to take on a new role or start a novel program may be crucial in influencing a person to seek out continuing education opportunities, whereas a person who is most concerned with the credentialing aspects of educational programs may be more likely to choose a degree program as a medium for personal advancement.

5) Specificity of purpose. Whereas a student first entering a professional program, like a child just entering school, is preparing broadly for a future which is relatively unknown, a continuing education student in a profession is likely to have a specific goal or set of goals; in a health field such a goal may be to increase clinical knowledge and ability relative to a specific patient population or relative to a specific class of patient problems. While there is an element of the unknowable in everyone's future, it is less pervasive to the mature, experienced person than it is to one who is in the very early stages of a professional career. The focus of the continuing education student is more specific than that of the basic professional student--perhaps emphasizing depth in a specific specialty.

Related to these characteristics which distinguish students enrolled in continuing education courses from those in other types of programs are complementary characteristics of the program themselves.

Program characteristics:

1) Corresponding to the specificity of purpose among mature students is a specificity of focus in program design. For example, a course may concentrate on a relatively narrow area such as care of the burn patient or oncology nursing. 2) Correspondingly, the length of such programs is limited. While full-time degree programs in nursing may run from one year (an unusual minimum) to four or even five years in length, it would be rare indeed for a continuing education course in nursing to exceed one year in length, even if classes are held only on a weekly, bi-weekly or monthly basis. If a full-time schedule of classes and clinical work is employed, a course running six weeks would be considered lengthy by most potential students. At the other end of the spectrum, a workshop of one day duration or even a single lecture or discussion may be considered to be a continuing education program as conceptualized in this paper. 3) Corresponding to other student characteristics--age, experience level, adult role, importance of intrinsic sources of motivation--is a third program characteristic. In degree programs faculty power to control student behavior is greatly augmented by the evaluative and credentialing functions of faculty members. In continuing education programs such sources of coercive power are generally absent; instructors must rely on charisma, displays of competence, and interpersonal influence in their efforts to have an effect on the lives, actions, and clinical performance of their students. 4) Finally, continuing education programs are frequently conducted by educational and service institutions which do not regard continuing education as a primary component of the institutional mission. Instead, a variety of pragmatic motives may be influential in causing an institution to take on

this role. A professional school within a university may undertake such a program as a money-making venture (whether funds are derived through student fees or through a government grant) or as part of a community relations effort. A hospital or other clinical agency may initiate programs designed to meet specific patient care needs noted within the institution. In situations in which it is basically marginal to the principal interests of the institution and its decision-makers, a continuing education program may receive a less than equitable share of institutional resources and support.

Evaluation Criteria

In the evaluation of continuing education programs, there are four levels of criteria which may be utilized; these relate to 1) participants' perceptions and opinions about the course 2) participants' knowledge and attitudes 3) participants' professional behaviors in their actual clinical work and 4) impact on patient status.

1) Perception/opinion data. Determination of student feelings about a course is an extremely common evaluation approach. Relevant data is likely to be collected via a questionnaire distributed at the end of the course; this questionnaire is designed to obtain information concerning students' opinions of the value of the course to them, in addition to obtaining specific perceptions concerning a number of basic course characteristics--areas such as organization, teaching quality, and difficulty level might be touched on. "To what extent did this course fulfill your objectives?" "What aspects of this course should be changed?" These are typical questions for a form of this kind. Basically, this approach is used to learn how satisfied participants are with the course and whether course characteristics as perceived by students are in accord with standards of good educational practice.

This approach may be criticized as yielding data which are soft--data which represent "subjective opinion" rather than "objective fact"--and there is some basis for this criticism. Naftulin, Ware, and Donnelly (1973) report a study in which an actor posed as a professor and delivered a lecture described as consisting of double-talk and circular arguments. The actor, however, presented his speech in a humorous and charismatic manner. The audience--professionals in the mental health field and in education--gave highly favorable ratings to this talk on an evaluation questionnaire. More recent studies have replicated this finding among college student samples, and they have begun the process of identifying key instructional and student variables involved in this phenomenon (Ware and Williams, 1975; Williams and Ware, 1975). The relevance of these findings to continuing education in clinical areas is underlined by two studies of programs in the medical field (McGuire, Hurley, Babbott and Butterworth, 1964; Williamson, Alexander and Miller, 1967); in these, although written subjective evaluation was favorable, objective measures of course effectiveness revealed that program goals concerning changes in the clinical practice of participants had not been achieved.

In spite of these results, the opinions and perceptions of participants in continuing education programs cannot be discounted. Such programs serve populations composed of professionals and adults. Decisions about program revision and program support are generally based largely on opinion; usually, the opinions of faculty and administrators are primary. However, experienced students who come into a program with specific, reality-based ideas of what they wish to learn may well be among the most astute judges of the value of a program. This is in sharp contrast to the

situation of younger, pre-professional students, who may have little practical knowledge of how the material they are learning will be utilized in their clinical practice; such students would seem to be less well qualified to make valid, independent judgments concerning the worth and potential application of an educational experience.

So, instead of dismissing opinion data as necessarily invalid, it should be used with caution. It is important to be aware of problems which may arise with its use and to be somewhat skeptical in one's interpretation of results. In addition, there are several techniques through which the validity of data of this kind can be increased. Low inference items--items which emphasize student perceptions in relation to previously established criteria rather than those which allow more room for student opinion--may be heavily relied on. For example, on a standard form used to evaluate each of a series of lectures, one may replace the opinion item, "Is the topic relevant to your nursing practice?" with the perception item "Were specific examples of nursing implications discussed?" However, the use of low inference items of this kind may be less important in continuing education programs than it is in other types of programs in which the judgments of students would be less highly trusted.

Another technique for increasing validity of opinion-perception data involves the use of a long-range data collection schedule. Six months after completion of a course, a person's perception of its value may be more accurate and more reality-based than it is on the final day of the course. After six months the student can report which aspects of his learning he has actually applied in his clinical practice. Still a third validity technique may be employed; this involves study of the patterns of relationships between opinion data and other categories of evaluation evidence. To the extent that moderate or substantial correlations can be demonstrated, the validity of all classes of data employed is made more conclusive.

2) Knowledge and attitude data. Typically, behavioral objectives of a course specify particular changes in knowledge or attitudes which are expected to occur as a result of the course. Evaluation is based on paper-and-pencil measures (knowledge examination, attitude scales) given at the start of the course and at the end. This methodology is designed to answer the following questions: Did knowledge increase as a result of the course? Did the desired attitude change occur? As with subjective data, a longitudinal approach to data collection adds to validity of the evaluation; one can ask not only "Did change occur?", but also "Is it retained?" Attitude change or knowledge gain which can be measured at the end of a program, but which is lost in the following weeks or months, is not likely to be of much value relative to the purposes of the course.

Generally, the measurement of knowledge via paper-and-pencil tests represents the most predominant of educational measurement techniques; it is certainly used more often than all other measurement techniques combined, especially in the process of determining the progress of individual students and assigning grades. But in prediction of success in non-school activities, current grading techniques are of questionnable validity. In several studies of the relationship between grades in medical school and post-graduate career success (as measured by a wide variety of criteria), it has been found that grades only predict grades; they do not predict out-of-school success (Loughmiller, Ellison, Taylor and Price, 1970; Taylor, Price, Richards and Jacobsen, 1964; Taylor, Price, Richards and Jacobsen, 1965). However, grading criteria increase in validity as the similarity in these criteria and in those used in later

judgments of performance become more similar. In many types of educational programs, opportunities for real world application are limited or nonexistent; so practical alternatives to measurement of knowledge through written examination are not readily available. This, however, is not the case among continuing education students; these students, for the most part, do function in real-world situations. Thus, the use of the examination as a basic evaluation technique is even less appropriate for those students than for those involved in other types of programs. What they know is less important than what they do--how they use what they know.

Similarly, attitude change measures may yield results of little validity relative to the purposes of a course. While experience and education in a field does not necessarily lead to adoption of a specific set of attitudes, it almost certainly will lead to knowledge about what attitudes are considered appropriate by the leading experts in the field. And most attitude instruments are very easy to fake!

3) Clinical process. Another approach to program evaluation in continuing professional education is to look for changes, not in knowledge or expressed attitudes, but in the performance which occurs in clinical situations. Structured observation of performance in specific clinical situations may be the most valid method of measurement in this area. However, practical considerations would support the use of rating scales (perhaps administered to participants and their supervisors or to selected peers or patients on a pre-post basis). Other written records of clinical actions might also be used, or participants might be asked directly such questions as "What things do you do differently now?" or "What programs have you initiated?" Again, a longitudinal data collection schedule is indicated, so that "sleeper effects" (changes which occur after a lapse of time) can be detected, and so that effects which are temporary can be distinquished from those which are permanent.

4) Patient impact. Finally the effectiveness of a continuing education program in health fields might be demonstrated through examination of patient outcomes which are related to the course. Here the evaluation question would be as follows: Does the course lead to an improvement in patient care as evidenced by improved patient condition? Measurement in this area might be based on chart reviews or other written records. However, direct measurement of patient outcome through independent observation and assessment may be preferable, although, in many cases, severe feasibility obstacles would first have to be overcome. While the use of this sort of criteria goes well beyond the conventional boundaries of the typical educational evaluation study, in evaluation of continuing education for practicing clinicians, it may represent the best possible operationalization of program goals.

Relationship Among Evaluation Criteria

Continuing education programs in the health professions, frequently claim an ultimate goal which is both ambitious and unambiguous--the improvement of patient care. This goal maybe oriented to some specific area of practice (i.e., oncology nursing) rather than to patient care in general; however, the need for justification of establishment of a continuing education course in terms of its potential for impacting on health care delivery system seems to be commonly accepted. The decision as to which evaluation criteria is most directly related to such a goal is equally clear. One's success in bringing about better patient care can best be measured by studying differences in patient care before and after the course and by determining

the extent to which discrepancies from the ideal exist in the two situations. While the other evaluation criteria provide information which may be useful or interesting they are further away from the basic question to be answered--whether the course leads to improved patient care as demonstrated by improved patient outcomes.

Several authors in the fields of educational evaluation, program evaluation, and clinical practice research have advocated the specification of relationships between process and product variables in order to build a theoretical causal chain in which the various program components are connected (Provus, 1972; Bloch, 1975; Suchman, 1967). Such a chain can then provide a basis for the determination of evaluation criteria.

For example, Bloch poses a nursing care problem in which a diabetic patient is in a negative health state because of failure to comply with an appropriate dietary regimen. This failure to comply is, in turn, caused by a lack of cognitive knowledge concerning dietary principles effecting diabetes. Bloch labels this sort of analysis a "diagnostic chain," and comments that although knowledge may be considered to be a valid outcome criteria, the optional evaluation plan would involve testing each link in the diagnostic chain.

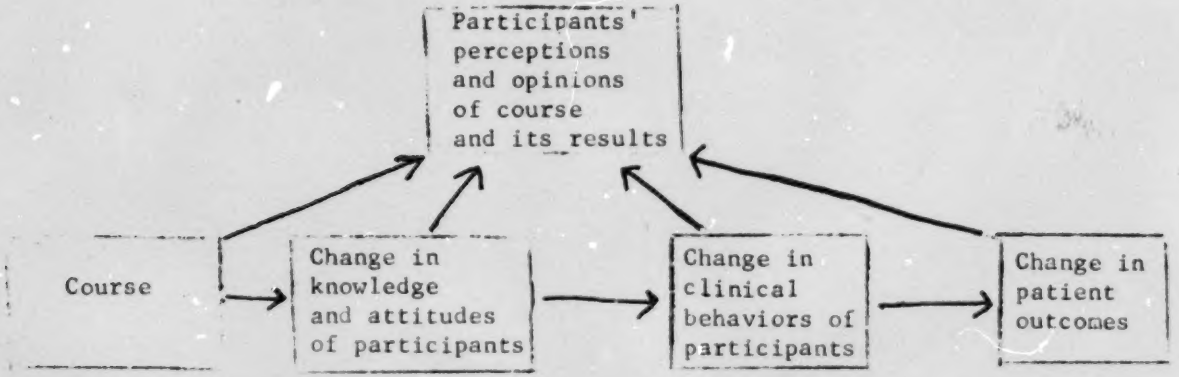
In discussing the relationships which may exist between immediate, intermediate, and ultimate objectives, Suchman introduces the concept of validity assumptions. The connection between lower level goals (i.e., to change physician's attitudes) and higher level goals (i.e., to improve patient care) can be made in one of two ways--1) by <u>proving</u> that achievement of intervening objectives is instrumental in achievement of the ultimate objective or 2) by <u>assuming</u> the validity of some expected relationships when full proof is lacking. In the absence of hard evidence that the extent to which a physician has a hopeful attitude toward cancer effects specific patient outcomes (perhaps patient satisfaction or amount of pain medication required), the use of physician hopefulness as a criterion variable involves a validity assumption--if the ultimate program goal is to improve patient care. However, if a relationship between physician hopefulness and patient outcome is shown either in the evaluation study or through reliance on previous research, no validity assumptions are necessary.

Weiss (1972) has pointed out that by testing the various links in a theoretical causal chain, a failure to achieve the ultimate objective may be attributed to a specific weak link in the chain. Understanding of the causes of program failure is much more useful to program developers than is the mere knowledge that ultimate objectives have not been achieved.

Returning to the classification of criteria useful for evaluation of continuing education programs in the health professions, it is possible to arrange three of the four criteria into a direct causal order--representing a generalized chain of events through which a successful course may impact on the health care delivery system. This is illustrated in Figure 1. If patient outcomes are to be improved, this can only be the result of changes in the clinician's behavior in relation to the patient's treatment or care. Such overt behavioral changes would seem to result from more subtle cognitive and affective changes which may, perhaps, be detected behaviorally through the use of written knowledge examinations and attitude scales. If such changes are detected between pre and post measurements (relative to a continuing

Figure 1

Generalized model of causal chain of relationships between possible evaluation criteria of a continuing education course in the health professions.


Participants' perceptions and opinions of course and its results
Course
Change in knowledge and attitudes of participants
Change in clinical behaviors of participants
Change in patient outcomes

education course), it seems logical to conclude that the course contributed to the change, and thus, set in motion a chain of events leading to improvements in patient care.

The placement of student perception data within this causal chain is a more complicated problem, as such perceptions can be influenced by any of the other links of the chain. In fact, it is appropriate to query students about their perceptions of the value of the course in relation to any of the other areas. For example, suppose that six months after course completion a physical therapist confronts a situation for which the course has prepared him (in terms of changes in knowledge or attitudinal base), he handles the situation in manner different than he would have previous to his participation in the course, and he observes that his action leads to a favorable patient outcome. This observation may increase the extent to which he perceives the course as worthwhile. In comparison to this sort of perception data, student opinion about course process (i.e., organization of course, difficulty of course) is shallow data indeed.

The distinction between process evaluation and product evaluation is well known among health education researchers. While product evaluation involves looking at the outcome of an educational program (perhaps in terms of student test scores or other indicators of student achievement), in process evaluation variables involved in the teaching-learning process are explored. However, this sort of separation between process and product variables may not be applicable to a continuing education program designed to impact on the health care delivery system. One expected <u>product</u> of such a course may be a change in the application of clinical <u>process</u> behavior employed by course participants. Thus, evaluation criteria used may be conceptualized as either product or process variables depending on one's frame of reference--with some variables (those referring to change in the practitioner) simultaneously representing both product variables (when conceptualizing evaluation of the educational system) and process variables (when conceptualizing evaluation of the clinical system).

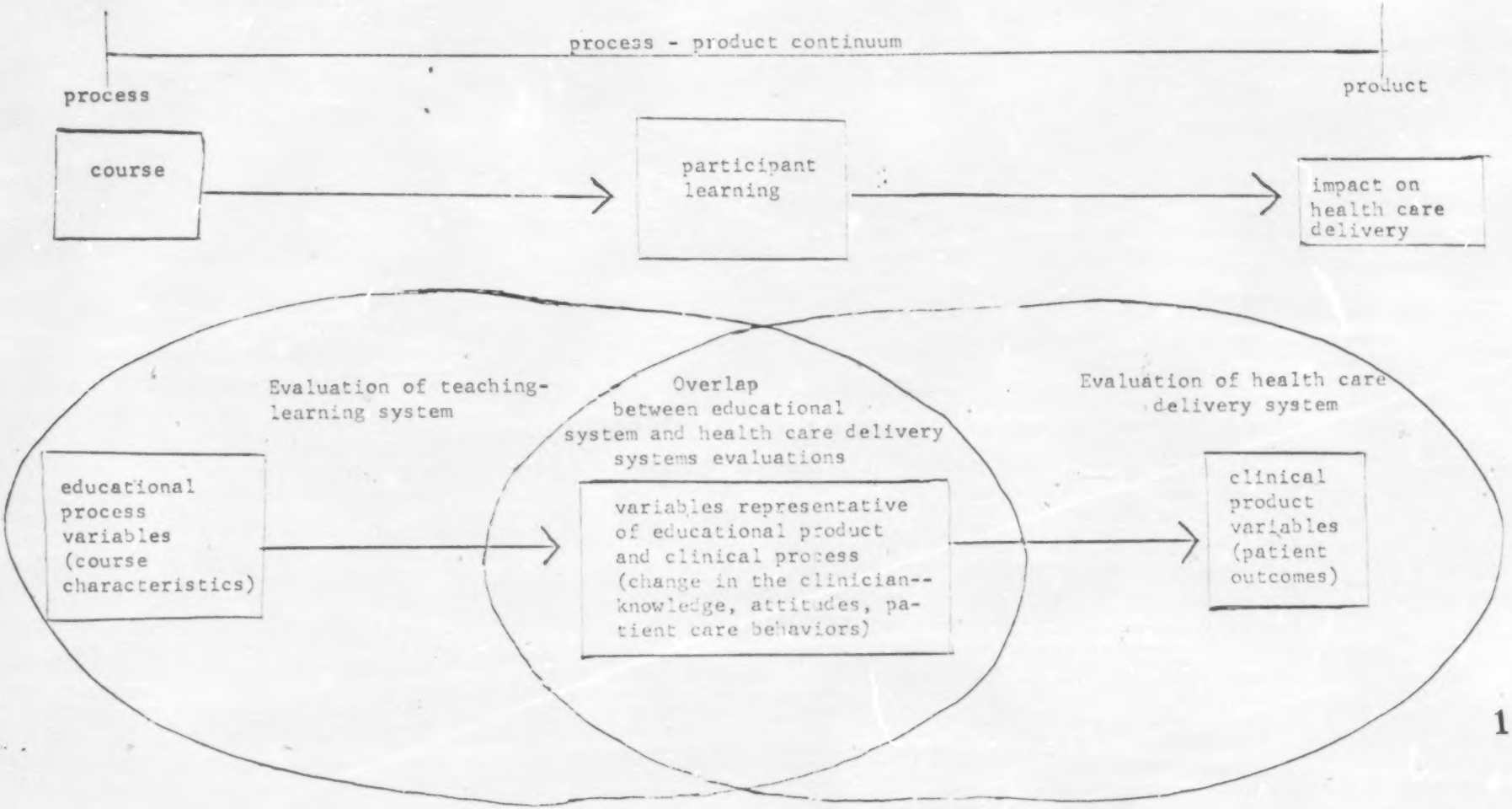
Looking at the course and its effects as a single entity, perhaps the most useful conceptualization is that of a process-product continuum, in which any single intermediary point may be conceived of as representing a product in relation to some criteria, and as representing a process in relation to others.

For example, as illustrated in Figure 2, participant learning (broadly operationalized as increased skills, increased knowledge or attitude change) would be considered an outcome variable in an educational evaluation study which concentrates exclusively on the teaching-learning system. However, in a broader evaluation which attempts to assess impact of an educational program on the health care delivery system (or some aspect of that system), participant learning would serve as an intervening or process variable.

Examined from this viewpoint, learning which occurs as part of the teaching-learning process should form a natural limitation on the potential impact of a program. (This limitation is approximate, not exact, as there is a portion of the potential impact which is independent of learning; i.e., trainee may be treated differently and thus have expanded opportunity to be innovative and do high quality work as a result of his supervisor's or his peer's awareness of his participation in

Figure 2

Process-product continuum in impact evaluation as related to overlap between evaluation of the teaching-learning system and of the health care delivery system


process - product continuum
process
product
course
participant learning
impact on health care delivery
Evaluation of teaching-learning system
Overlap between educational system and health care delivery systems evaluations
Evaluation of health care delivery system
educational process variables (course characteristics)
variables representative of educational product and clinical process (change in the clinician--knowledge, attitudes, patient care behaviors)
clinical product variables (patient outcomes)

the program). However, such situations aside, it may be generally assumed that the greater is degree of learning which has occured, the greater is the potential for quality of patient care to be effected--though this potential may be filled to a greater or lesser extent.

Literature Review

A review of the literature of evaluation of continuing education in several health professions reveals that although reports frequently begin with a statement that improvement in patient care is the program goal, the data presented tends to focus on the opinions, knowledge, attitudes or behavior of the clinician rather than on the potential product of these phenomena--patient outcomes. This tends to be true in nursing (Popiel, 1971; Popiel, 1973; McMahon and Neuman, 1972; Parker, Pierce and Sturm, 1975; Lynch, 1972; Barlow and Bruhn, 1973; Rudkin and Hamister, 1973; Heffernan, 1972), in medicine (Long, 1969; Ware, Donnelly and Naftulin, 1970; Donnelly, Ware, Wolken and Naftulin, 1972) and in dentistry. Yet, in each of these fields, several model studies may be identified. The methodologies used in these studies should be examined for possible emulation by others, and the findings derived may be utilized in helping us to assess where we should go from here.

Among published evaluation studies of continuing education in nursing, perhaps the work reported by Condon (1971; 1972) comes closest to the conceptualization advocated in this paper. She has described a program on teaching methods designed for hospital personnel involved in in-service education. Rather than basing evaluation on characteristics of program participants, she focused on the product of change in the participants. To the extent that the participants have gained competence in teaching elements of nursing process, this should become apparent thru study of changes in the use of nursing process within their respective hospitals. She reported data, submitted by head nurses, concerning the consistency of performance of activities related to each aspect of nursing process, and, on the basis of this data, leads the reader to the conclusion that a program on teaching methods for hospital in-service personnel was responsible for improvement in quality of nursing among participating hospitals. Unfortunately, the final step--demonstrating that the increased use of nursing process activities led to improved patient outcomes--was not taken.

In one current effort being conducted at Wayne State University, an attempt is being made to evaluate an oncology nursing program by means of QualPac (Wandelt and Ager, 1970), which is a system for determining the quality of nursing care received by a patient. However, this effort has met with numerous obstacles and whether it can be successfully implemented remains yet to be seen (Jansen, 1977).

In dentistry, one researcher (Chambers, 1976; Chambers and Hamilton, 1975) has undertaken a line of studies designed to determine how practicing dentists learn. He has reported that the goals of practitioners taking course are quite discrepant with those of program planners. While 63% reported being generally interested in the topic as a motive for registering, only 5% reported as a motive wanting to change their practice. In addition, concerning a course in four-handed, sit-down dentistry, evaluation data revealed that there were pre-post changes in behavior of dentists (reported by trained observers), but that the dentist themselves were poor at evaluating their own learning.

A number of researchers in the field of medical education have gone beyond self-report and supervisor report to audit or observational methods in looking at changes in clinical process variables in the actual practice situation. However, most of these stop short of examining patient outcome. A series of lectures concerning the value of Papanicolaou smears in detecting cervical cancers was evaluated through use of a formula in which the number of Pap smears performed by the physician was divided by the number of patient visits by females over age 20 (Kane and Bailey, 1971). Instruction in identification of unknown heart sounds was evaluated, in part, through a chart review in which quality of cardiovascular examination as recorded was determined (McGuire, Hurley, Babbott and Butterworth, 1964). A workshop designed to increase physician response to unexpected abnormalities was also evaluated through examination of patient records (Williamson, Alexander and Miller, 1967). These studies are consistent in reporting that the programs evaluated were ineffective in changing how physicians practice, although the use of other types of evaluation data (satisfaction, knowledge) yielded more positive results. These conclusions underline the difficulty in achieving success if change in clinical behavior is the principle evaluation criterion.

In two similar, but broader studies, a number of process criteria were measured. Rubenstein (1973) evaluated a Back-to-Medical-School program for practicing physicians by looking at utilization rates of a variety of techniques during time intervals before and after lectures during which the techniques were either condemned or advocated. Techniques of interest included both treatment procedures (i.e., number of transfusions of whole blood and number of transfusions of packed cells) and diagnostic techniques (i.e., lung scan). Data was collected through review of hospital records. In the other study (Shepard, 1971), a program designed to promote use of intensive care techniques among physicians from small rural hospitals was evaluated by assigning an instructor/observer to work with the physician--participant in his own hospital. One of the tasks of the observer was to record utilization or non-utilization of specific techniques in cases for which they were indicated. One other study, in which evaluation of a short workshop in Tonometry was reported, the number of tonometric exams conducted by a sample of the participants was used as one criterion of success (Kaplan, 1974).

Two medically-based studies, however challenge conventional notions about feasibility in evaluation studies; in these, not only physician behavior, but also patient outcomes are considered. One was based on the epidemiological model for research. Lewis and Hassanein (1970) evaluated a ten year series of programs organized by the University of Kansas by associating rates of attendance with maternal death rates, perinatal death rates and surgery rates between the eleven regions of the state. While the design of this study is creative, it leaves one glaring deficiency--physicians in a region may choose to study those topics which are most problematic to the region. So no inferences about cause and effect can be drawn from this data.

Another creative study was conducted by Inui, Yourtee and Williamson (1976). An experimental design was used to evaluate the use of physician tutorials concerning hypertension treatment. In addition to measures of physician behavior (time spent on physical examination, history taking, and patient education; and topics of chart notations), several patient outcome measures were used (patient knowledge, patient compliance, blood pressure readings).

Many evaluation studies concerning continuing professional education are conducted on a small scale and are intended for institutional or agency use only.

A survey of nursing continuing education programs in the Midwest (Forni and Overman, 1974) has shed some light on the nature of a sample of unpublished evaluation studies. Data on 57 continuing education programs was obtained and analyzed; of these programs 77% reported that some sort of written evaluation form was used, and almost half claimed to have other evaluation information as well (class projects, evidence of changed behavior, increased enrollment in subsequent courses and verbal or written feedback). Follow-up studies of skills or attitudes were conducted by 11% and 14% of the programs, respectively. Of the 18 projects for which evaluation material was submitted, the surveyors concluded that for 14 this material consisted of a happiness index form. For three others an examination was included. Though six types of impact information collected were listed by the authors, these, as in the published studies, focused on the nurse-learner, not on the patient recipient of care.

It is clear that the conventional wisdom of educational evaluation has had an influence on the conduct of the evaluations described. Seventy-two percent claimed to have constructed objectives which were in measurable form; that development of such objectives is important is a key message among writers in the field of educational evaluation.

The influence of writings of educational researchers may not be an entirely positive one, however; to the extent that evaluations of continuing education programs in health fields are based on typical methodological considerations appropriate to the evaluation of youth education, the distinctive aspects both of continuing education and of education for clinical practice may be ignored. While the case for continuing education in the health professions must be made in terms of its impact on patient outcomes, most evaluation data remains one or more steps removed from this criteria. To the extent the conclusive evidence demonstrating the links between perceptions, attitudes, knowledge, clinical process and patient outcomes is lacking, evaluation which is based on use of such participant or process variables is dependent on validity assumptions of which the accuracy may be open to question.

## Limitations of Evaluation

From the analysis presented in this paper, one might conclude that evaluations of continuing education programs in the clinical professions must deal with the effect of the course on patient outcomes; if they do not, the process of evaluation is not complete. However, being a working evaluator, the present author recognizes that the expectation that measurement of patient outcomes will become a standard component of evaluations of education programs is totally unrealistic. Rather than improving the quality of evaluation, emphasis on the need for such measurement would probably have the effect of convincing practitioners that good evaluation is impossible, and therefore, one should not bother to attempt it.

The obstacles to routine measurement of patient outcome data as a technique for evaluating educational programs are probably insurmountable--at least at our present level of knowledge. Varying goals and settings of individual course participants, geographical spread of course participants, discrepancies in record keeping systems of various agencies, interdependence of health professionals, and lack of appropriate,

reliable, and valid outcome measures for evaluating quality of care in some areas would be among the problems confronted. The need to avoid excessive interference with program planning must also be a concern. Program evaluation, unlike other forms of quality of care research, is most conclusive if a pre-post design is employed. The use of such a design in combination with patient outcome criteria must, however, place restrictions on registration procedures of a program; advance registration would be required. If this causes a reduced program enrollment, it would, in all probability, actually reduce the cost-effectiveness of the program. If the primary purpose of the program is to educate, rather than to evaluate, this is probably an unacceptable price to pay.

We now come to the crux of the problem. The type of evaluation discussed in this paper is most often undertaken as an activity of secondary importance. It may be referred to as "tacked on" evaluation, because it is a tacked on project component--tacked on, perhaps, as an afterthought, and perhaps as a response to the requirements of some funding agency. Often, evaluation planning is not begun until other aspects of program design are well formulated--perhaps even implemented. The energy and resources available for evaluation of the tacked on variety is likely to be some small proportion of that available to the entire project; five or ten percent would be typical figures.

Conclusions--Measurement, Inference and Supporting Research

Given these obstacles and limitations, if evaluation of continuing education programs is to be both successful and realistic, it cannot stand alone; instead, it should be supplemented by two other lines of research. 1) All research into the effects of continuing education does not need to be of the "tacked on evaluation" variety. It is possible to create projects of which the primary purpose is to study continuing education, rather than to practice it. Such studies may serve as demonstration projects making strong inroads into the solution of feasibility obstacles to measurement of patient outcomes. Further, studies of this kind may be extremely useful in creating knowledge about relationships between educational processes experienced by clinicians, the resulting state of patient well-being, and the intervening variables which mediate this association. Research into factors which may influence the extent of change in clinical behaviors is also sorely needed, given the findings of some studies that behavioral change was not achieved, despite knowledge gain and a feeling that the program was valuable. These tasks are best reserved for basic research into the nature of continuing education. These are not functions which can be performed routinely in the evaluation of educational programs. There are too many obstacles imposed by the real world situation. For routine evaluation to establish the presence or absence of behavioral change (or perceptions of such change) is challenge enough if one is to keep the commitment and resources devoted to evaluation within reasonable limits relative to that of the program being evaluated.

2) The second line of research on which continuing education evaluators must rely falls within the realm of clinical practice research. In general, to show that specific behavior inputs by health professionals are related to specific patient outcomes is the task, not of program evaluation, but of clinical practice research. Experimental studies, not tacked-on evaluations, are needed for this purpose; the results of such experiments can then be used by planners and evaluators of educational programs in drawing inferences about possible program results and in minimizing the

need for measurement in some areas.

Being limited in resources and facing powerful practical challenges, evaluation studies must be limited in scope. Rather than attempting to measure every key variable of the process-product continuum, a balanced combination of measurement and inference may be employed. Selectivity must be exercised in choosing variables for measurement. Gaps in measurements taken may be filled by drawing inferences concerning relationships between variables which are measured and those which are not. Evidence to support such inferences may be obtained from a number of sources. Research studies in the areas cited above may be referred to in documenting an inferred relationship between for example, a particular clinician behavior and a patient outcome or a relationship between attitude change of the worker and a change in his clinical behaviors.

If appropriate research can not be located, consensus of expert opinion may also serve as a basis of inference. For example, in the evaluation of one component of a continuing education program in oncology nursing, a scale for the measurement of clinical behavior was developed through a content validity study; a sample of national experts rated the importance to effectiveness in cancer nursing of each of a long series of behaviors (i.e., assesses physiological strengths and weaknesses of the cancer patient). They were also invited to add other areas which should be included. A more simple procedure of citing references from the clinical literature may also be employed to support an inference through the use of expert opinion.

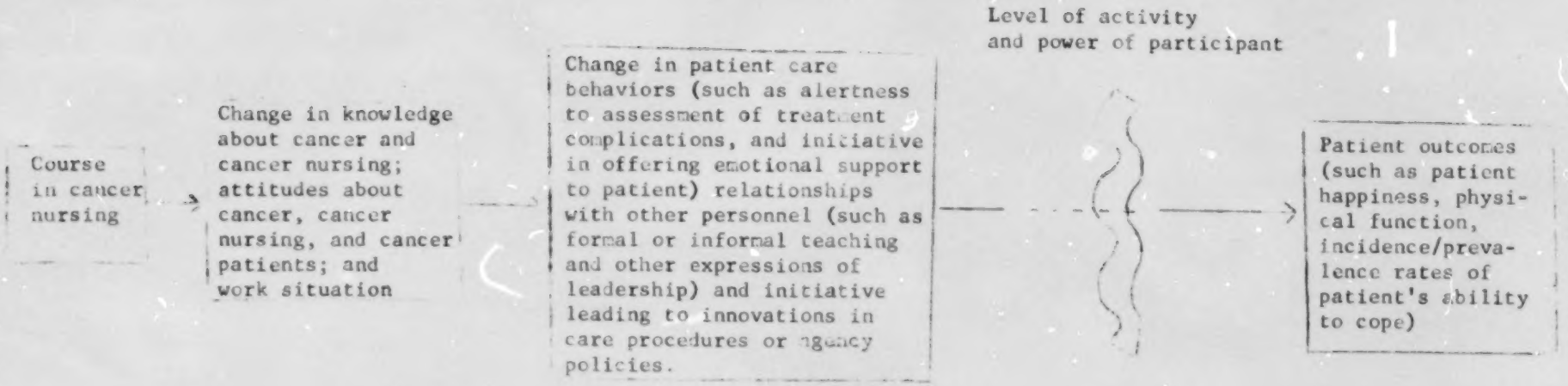
Another source of support for inferences may come from the evaluation study itself. If one postulates that clinical change will be the result of participants' knowledge gain, but only the former attribute is measured, the finding that such change was achieved would enable the researcher to effectively infer that knowledge gain was also achieved.

In developing an evaluation study in which a balance of measurement and inference is used in determining program effects, it is necessary to be completely clear about the nature of the inferences being made, the documentation on which these inferences are based, and the relationships between these inferences and the measurements which are taken. The contrast between conclusions based on measurement and those based on assumption must be made explicit. One method of accomplishing this is through the construction of a specific program model--a model of the expected chain of events through which a specific educational program may influence patient outcomes. In most cases, such a model would closely parallel the generalized model of relationships between criteria presented earlier in this paper.

An example from an evaluation currently underway may serve to illustrate this point. The oncology nursing course referred to earlier is one of a number throughout the country being sponsered by the National Cancer Institute of the Department of Health, Education and Welfare with the goal of improving the quality of nursing care available to cancer patients. Figure 3 illustrates the chain of events expected to result from this course. It is anticipated that this course will increase participants' knowledge of cancer nursing and cause some attitude change. It may also effect their work situation in some way by, perhaps, increasing their prestige or power or by causing their contact with cancer patients to increase. Each of these changes can, in turn, contribute to a change in the clinical performance of the nurse. Direct patient care variables, leadership among clinicians, and initiative in promoting innovations on the unit are key areas within this category of events. The extent to which these

Figure 3

Specific chain of events expected in relation to a continuing education course in oncology nursing.


Level of activity
and power of participant
Course in cancer nursing
Change in knowledge about cancer and cancer nursing; attitudes about cancer, cancer nursing, and cancer patients; and work situation
Change in patient care behaviors (such as alertness to assessment of treatment complications, and initiative in offering emotional support to patient) relationships with other personnel (such as formal or informal teaching and other expressions of leadership) and initiative leading to innovations in care procedures or agency policies.
Patient outcomes (such as patient happiness, physical function, incidence/prevalence rates of patient's ability to cope)

behaviors lead to changes in patient outcomes (such as, perhaps, patient happiness or incidence of infection) is mediated by the nurse's activity level. Other things being equal, a person working full time will have a greater impact than one who works part time, and one in a position of authority over others will have greater influence than someone who occupies a lower spot in the institutional pecking order. (For the most part, these activity variables are characteristics of the person, not of the course; while it is possible for the course to have some influence on these, usually this influence would not be of great magnitude.)

Finally, the effect of clinical behavior change on patient outcomes is dependent on the nature of the clinical behavior change. To the extent that research and clinical observation have established the effectiveness of similar changes in other situations, one may be on fairly solid ground in inferring that the changes produced by the course will lead to improved patient outcomes. To the extent that the potential consequences of these changes are unknown, the impact of the course on patient care remains in doubt until that impact is measured--regardless of the course's effectiveness in changing the clinical behaviors of its participants.

While such a model may seem to be a simplistic representation of a much more complicated reality (and indeed any real working model would need to be more complex), the utilization of such a model clarifies the consequences of certain evaluation criteria on which data collection will be concentrated.

Depending on the purpose of evaluation, either of two sets of principles may be followed in making such decisions. 1) In summative evaluation--determination of the worth of a program--one's primary interest is in the last point in the chain of events. The goal of this type of evaluation is to draw conclusions about the impact of the course on the welfare and satisfaction of patients; one may maximize the certainty with which such conclusions may be presented through two techniques:

a) One may minimize the degree of inference and assumption required by taking measurements at the last class of events at which such measurements may feasibly be taken.

b) One may support the validity of inferences which are made through reference to previous research, expert opinion or other documentation.

2) In formative evaluation--evaluation which is orientated to identifying program deficiencies so that improvements can be made--measurement of the end point in the chain is not sufficient, especially if negative (or partially negative) results are obtained. While positive findings at any evaluation criteria level tend to validate the accuracy of the postulated chain of events to that point, negative results on latter events provide little information on which to base program revisions. That is, if clinical behavior change occurs it is safe to assume that change in knowledge, attitude or work situation also occurred--even if direct measurements designed to verify these assumptions are not obtained. However, if no clinical behavior change occurs, identification of the weak link of the chain of events model requires that measurement of earlier stages in the process are available. Did the course fail to produce knowledge gain? Or did achieved knowledge gain fail to lead to change in clinical behaviors? If one wishes to make improvements in a not totally effective program, this sort of information is essential. Thus, for effective formative evaluation, it is necessary to take measurements at various stages in the chain of events

model, though for summative evaluation it is not.

In the oncology nursing course discussed above, both formative and summative evaluation are of interest. A decision not to tackle the problems of patient outcome measurement was made, so change in clinical behaviors has become a major point of focus; this is measured through use of a behavior rating scale completed by course participants and by their supervisors both before the course and at various time intervals following the course. More open-ended questionnaires and critical incidents forms are other sources of data about clinical behavior change. In order to serve the needs of formative evaluation, perceptions of the course and measurements of knowledge gain and attitude change are also obtained--but in dealing with this data emphasis is on providing early feedback to project planners, not on establishing reliability and validity of the forms nor on accumulating data for a major statistical analysis.

It is the feeling of this writer that use of the line of analysis presented in this paper will result, most often, in choice of variables representing change in clinical behavior as the primary evaluation criteria. While the areas of knowledge gain and attitude change should not be ignored, these are an extra step removed from the ultimate program goal--if that goal is to improve patient outcomes. Success in these areas is a long way from program success, and basing a conclusion that the program is effective on such data requires a giant inferential leap. Deriving a conclusion that patient care is improved from data concerning clinical behavior of the practitioner also requires an inferential leap, but this a leap of more manageable proportions, and the magnitude of the leap reflects the state of knowledge in that area of clinical practice with which the program is concerned; thus, the validity of the evaluation is limited only by gaps in the validity of clinical practice. The responsibility for linking clinical behaviors and patient outcomes falls primarily, not within the realm of educational program evaluation, but within the realm of clinical practice research.

Through a realistic conception of what evaluation may reasonably be expected to accomplish, a reliance on appropriate research being conducted in education and in the health fields, and a balanced, judicious combination of measurement and inference, the evaluation of continuing education programs may be expected to lead to improvements in our knowledge about this important educational format; it should provide insights upon which planning decisions may be made with a reasonable, though not total, level of certainty.

## REFERENCES


Barlow, D.J. & Bruhn, J.G. Role-plays on television--a new teaching technique. Nursing Outlook, 1973, 20, 265-266.

Bloch, D. Evaluation of nursing care in terms of process and outcome: issues in research and quality assurance. Nursing Research, 1975, 24, 256-263.

Caplan, R.M. Measuring the effectiveness of continuing medical education. Journal of Medical Education, 1973, 48, 1150-1152.

Chambers, D.W. How can learning experiences in continuing dental education be evaluated in relation to patient benefit? The Journal of the American College of Dentists, 1976, 43, 238-248.

Chambers, D.W. & Hamilton, D.L. Continuing dental education: reasonable answers to unreasonable questions. Journal of the American Dental Association, 1975, 90, 116-120.

Condon, S.M.B. Teaching the teachers. Nursing Outlook, 1971, 19, 804-806.

Condon, S.M.B. Inservice education...impact on patient care. Journal of Continuing Education in Nursing, 1972, 3, 34-38.

Donnelly, F.A., Ware, J.E., Wolkon, G.H., & Naftulin, D.H. Evaluation of weekend seminars for physicians. Journal of Medical Education, 1972, 47, 184-187.

Forni, P.R. & Overman, R.T. Does continuing education have an effect on the practice of nursing? a survey. Journal of Continuing Education in Nursing, 1974, 5, 44-51.

Heffernan, B.M. Keeping nurses up to date. Nursing Outlook, 1972, 20, 265-266.

Inui, T.S., Yourtee, E.L. & Williamson, J.W. Improved outcomes in hypertension after physician tutorials: a controlled trial. Annals of Internal Medicine, 1976, 84, 646-651.

Jansen, E. Description of evaluation model of oncology nursing project at Wayne State University. Zerox publication distributed at Oncology Nursing Program Evaluation Meeting, New Haven, Connecticut, 1977.

Kane, R. & Bailey, R.Q. Evaluation of a postgraduate educational programme in early cancer detection. British Journal of Medical Education, 1971, 5, 134-137.

Knowles, M. The adult learner: a neglected species. Houston: Gulf Publishing Company, 1973.

Kramer, M. Reality shock: why nurses leave nursing. Saint Louis: Mosby, 1974.

Lewis, C.E. & Hassanein, R.S. Continuing medical education - an epidemiological evaluation. New England Journal of Medicine, 1970, 282, 254-259.

Long, L.D. The evaluation of continuing education efforts. American Journal of Public Health, 1969, 59, 967-973.

Loughmiller, G.C., Ellison, R.L., Taylor, C.W., & Price, P.B. Predicting career performance of physicians using the biographical inventory approach. Proceedings, 78th Annual Convention, APA, 1970, 5 (Part. 1), 153-154.

Lynch, L. Toward appropriate change in behavior: measuring knowledge, attitude and skill in continuing education. Journal of Continuing Education in Nursing, 1972, 3, 6-10.

McGuire, C., Hurley, R.E., Babbott, D. & Butterworth, J.S. Auscultatory skill: gain and retention after intensive instruction. Journal of Medical Education, 1964, 39, 120-131.

McMahon, J. & Neuman, M.M. Tool for evaluating the impact of an in-service program on nursing care. Journal of Continuing Education in Nursing, 1972, 3, 5-7.

Naftulin, D.H., Ware, J.E. and Donnelly, F.A. The Doctor Fox lecture: a paradigm of educational seduction. Journal of Medical Education, 1973, 48, 630-635.

Parker, B., Pierce, M. & Sturm, I.E. Problem oriented record. Journal of Continuing Education in Nursing, 1975, 6, 35-39.

Popiel, E.S. Evaluation of the effectiveness of the continuing education course--"new perspectives in college health nursing." Journal of the American College Health Association, 1971, 19, 169-173.

Popiel, E.S. Long-range evaluation of a continuing education course. Journal of Continuing Education in Nursing, 1973, 4, 15-17.

Provus, M.M. The discrepancy evaluation model. In P.A. Taylor & D.M. Crowley (Eds.), Readings in curriculum evaluation. Dubuque, Iowa: Wm. C. Brown Company Publishers, 1972.

Rubenstein, E. Continuing medical education at Stanford: the back-to-medical-school program. Journal of Medical Education, 1973, 48, 911-918.

Rudkin, G.J. & Hamilter, S.I. Evaluation of a training program for nursing assistants. Nursing Outlook, 1973, 21, 404-407.

Shephard, R.M. Evaluation of continuing education. Proceedings of the Royal Society of Medicine, 1971, 64, 149-150.

Suchman, E.A. Evaluation research: principles and practice in public service and social action programs. New York: Russell Sage Foundation, 1967.

Taylor, C.W., Price, P.B., Richards, J.M., & Jacobsen, T.L. An investigation of the criterion problem for medical school faculty. Journal of Applied Psychology, 1964, 48, 294-301.

Taylor, C.W., Price, P.B., Richards, J.M. & Jacobsen, T.L. An investigation of the criterion problem for a group of medical general practitioners. Journal of Applied Psychology, 1965, 49, 399-406.

Wandelt, M.A. & Ager, J.W. Quality patient care scale. New York: Appleton-Century-Crofts, 1970.

Ware, J.E., Jr., Donnelly, F.A., & Naftulin, D.H. The evaluation of medical education. Journal of Medical Education, 1970, 45, 813.

Ware, J.E. & Williams, R.G. The Dr. Fox effect: a study of lecturer effectiveness and ratings of instruction. Journal of Medical Education, 1975, 50, 149-156.

Weiss, C.H. Evaluation research. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972.

Williams, R.G., & Ware, J.E. Validity of student ratings of instruction under different incentive conditions: a further study of the Dr. Fox effect. Paper presented at AERA Convention, 1975, Washington, D.C.

Williamson, J.W., Alexander, M., & Miller, G.E. Continuing education and patient care research: physician response to screening test results. Journal of the American Medical Association, 1967, 201, 118-122.